# Experiments with Synthetic Speech Concerning Quantity in Estonian

Ilse Lehiste

# Experiments with Synthetic Speech Concerning Quantity in Estonian

Ilse Lehiste

## 1. Introduction

This paper constitutes a first report on an experiment designed to test the relevance of various suprasegmental parameters in the perception of quantity in Estonian. The test materials consisted of synthetically produced acoustic stimuli, intended to sample systematically the acoustic spaces containing the minimal triples taba - tapa - tappa and sada - saada! - saada. The synthesis was performed by means of a Digital Data Processor (DDP 24) computer at the Bell Telephone Laboratories.[1] The synthesis was carried through entirely by rule, i.e. no attempt was made to imitate a known speaker. The stimuli will be described below in more detail. Test tapes containing randomized stimuli were presented to 26 listeners, who are native speakers of Estonian, at the Experimental Phonetics Laboratory of the Academy of Sciences in Tallinn, Estonia.[2] Two tapes were used, one for the taba - tapa - tappa set, the other for the sada - saada! - saada set; each contained 252 stimuli. As there were 26 listeners and each made 504 judgments, the data consist of 13,104 individual judgments. The statistical evaluation of the materials is in progress; however, some results are already available, and a preliminary survey is given below.

## 2. Taba - tapa - tappa

The synthetic material was designed to test the ranges of /p/ durations which would be assigned to the three quantities, and the contribution of second syllable duration to the perception of the three test words. The duration of /p/ was varied in twenty-one 10 msec steps over a continuous range from 40 to 240 msec. Each of the 21 /p/-durations was combined with three durations for the second vowel: 180 msec, 120 msec, and 90 msec. The duration of the first vowel was kept constant at 120 msec; the fundamental frequency was likewise constant (at 120 Hz). The total of 21 x 3 = 63 stimuli was arranged in four different randomizations and presented to listeners, who had to assign each stimulus to one of the three words taba, tapa or tappa. The listeners thus made a forced-choice linguistic judgment rather than a phonetic judgment. Each listener gave 252 responses, for a total of 6,552 responses. The results of the listening test are summarized on the following figures and tables.

Table 1 and Figure 1 show the general effect of second syllable duration on the assignment of the words to quantities one, two and three. It is obvious that a second syllable duration of 180 msec

favors assignment to quantities one and two: the number of taba and tapa responses is greatest under this condition. On the other hand, a second syllable duration of 90 msec favors assignment of the word to quantity three.

Tables 2-4 and Figures 2-4 show the number of judgments as taba, tapa or tappa as a function of the duration of intervocalic /p/. Each of the three tables and figures represents judgments associated with one of the three second syllable durations. The discussion of the tables and the figures will be limited to a few brief comments.

If we consider the crossing-points of curves representing taba, tapa, and tappa judgments as 'phoneme boundaries' between quantities 1, 2 and 3 of the intervocalic consonant, then we note that the phoneme boundary between /p/ in quantity 1 and /p/ in quantity 2 depends only slightly on the duration of the second vowel: with decreasing second syllable duration, the boundary shifts from approximately 110 msec for a second syllable duration of 180 msec to 105 msec for a second syllable of 120 msec, and to 100 msec for a second syllable of 90 msec. However, the boundary between quantities 2 and 3 appears crucially affected by the duration of the second syllable. Figure 2 shows that if the second syllable had a duration of 180 msec, the boundary between tapa and tappa was at 225 msec, and even with the longest duration, 240 msec, the differentiation between long /p/ and overlong /p/ was very tenuous. With second syllables of 120 and 90 msec, the boundary between long and overlong intervocalic /p/ occurred at 175 and 170 msec respectively.

### 3. Sada - saada! - saada

The set of test items designed to test the perception of quantity in disyllabic words of the type sada - saada! - saada is a little more complicated. This time there were three variables: duration of the vowel of the first syllable, duration of the vowel of the second syllable, and the fundamental frequency pattern distributed over the two syllables. The duration of the first vowel varied in seven 20-msec steps from 120 to 240 msec, while the duration of intervocalic /t/ was kept constant at 60 msec. Each of the first syllables was combined with the same three second syllable durations as in the previous case, namely 180 msec, 120 msec, and 90 msec. Furthermore, each disyllabic stimulus was synthesized with three fundamental frequency patterns: a level pattern (monotone at 120 Hz), a step-down pattern (with the first syllable level at 120 Hz and the second syllable level at 80 Hz), and a falling pattern (first syllable falling from 120 Hz to 80 Hz, second syllable level at 80 Hz). The total number of stimuli was again 7 x 3 x 3 = 63, the total number of items on the randomized tape was 252, and the number of judgments was 6,552.

The results are presented on Tables 5-8 and Figures 5-11. Again, only a few descriptive comments will be given this time.

Table 5 and Figure 5 show the influence of second syllable duration and fundamental frequency pattern on the overall classification of stimuli as sada, saada! and saada. As is apparent from the left

half of Figure 5, the influence of second syllable duration was comparable to what was observed with the set taba - tapa - tappa: a longer second syllable favored judgments for quantities 1 and 2, and disfavored judgments as quantity 3, while the shortest second syllable increased the number of quantity 3 judgments in a substantial manner.

This effect is, however, rather limited compared to the influence of the fundamental frequency pattern. As becomes apparent from Figure 5, the monotone condition was relatively neutral. The step-down pattern, with the first syllable level at 120 Hz and the second syllable level at 80 Hz, produced the greatest number of quantity 2 judgments and the smallest number of quantity 3 judgments. It is important here to notice that the step-down pattern actually decreased quantity 1 judgments; for quantity 1, the monotone pattern was the most favorable one.

Conversely, the falling pattern significantly increased the number of quantity 3 judgments and decreased quantity 2 judgments. This decrease took place almost exclusively at the expense of quantity 2, since the number of quantity 1 judgments remained practically constant.

The phoneme boundaries for the duration of the first vowel are rather difficult to establish, since both the second syllable duration and especially the fundamental frequency pattern have such a strong influence on perception. Some of the problems are illustrated on the figures.

Figure 6 shows the assignment of stimuli to quantities 1, 2 and 3 with a second syllable of 180 msec and with a level fundamental frequency pattern. It may be recalled that these two conditions favor assignments to quantity 1 and disfavor assignments to quantity 3. As is obvious from the figure, the overlap between quantities 1 and 2 occurs at approximately 160 msec, while the two curves representing quantities 2 and 3 do not overlap at all. Even at the longest duration, 240 msec, 73 out of 104 judgments were still made in favor of quantity 2.

Figure 7 shows the number of judgments with the same second syllable duration--180 msec--but with a falling fundamental frequency pattern on the first syllable. As was mentioned before, this pattern favors assignments to quantity 3 and disfavors assignments to quantity 2, leaving quantity 1 practically unaffected. The phoneme boundary between quantities 1 and 2 has shifted only very slightly, from 160 msec to approximately 155 msec. It is now also possible to talk about a phoneme boundary between quantities 2 and 3: it would fall at about 210 msec.

Figure 8 shows assignments to the three quantities with a short second syllable (90 msec) and monotone fundamental frequency. As may be remembered, the short second syllable favors assignments to quantity 3, while the monotone fundamental frequency pattern is relatively neutral. A characteristic of all three curves is the extensive overlap between them and the fact that all three curves peak at approximately 75%. The reliability of recognition here obviously was not very great; the phoneme boundaries, however, seem not to have been affected.

Figure 9 shows assignments to the three quantities under conditions maximally favoring quantity 3: a short second syllable (90 msec) and a falling fundamental frequency pattern. The reduction of the number of quantity 2 judgments is particularly striking: even at the 160 msec duration, which produced the greatest number of quantity 2 judgments, their number did not exceed 64 (out of 104). The phoneme boundary between quantities 1 and 2 is not affected, but the boundary between quantities 2 and 3 has now shifted from 210 to 175 msec. The peak of the curve has shifted from 180 msec with level fundamental frequency (Figure 8) to 160 msec.

Figures 10 and 11 summarize the influence of fundamental frequency patterns on assignment to quantities 2 and 3. The second syllable in these two sets of examples was constant at the most neutral, intermediate value, namely at 120 msec.

Figure 10 shows assignments to quantity 2. It is obvious that the left-hand slope of the curve depends very little on the fundamental frequency pattern: the phoneme boundary between quantities 1 and 2 is barely affected by the fundamental frequency. On the other hand, the position of the peak and the phoneme boundary of quantity 2 with regard to quantity 3 are both strongly affected: the peak shifts from about 210 msec with the step-down curve to 180 for the monotone and to 160 for the falling pattern.

The converse situation appears on Figure 11, which shows the influence of fundamental frequency on assignments to quantity 3. Here the neutral pattern produced the smallest number of assignments, the step-down pattern increased the number of quantity 3 judgments somewhat (although the curve never reached 70%), and the falling pattern both steepened the slope of the curve and made it reach a higher peak. It should be noted that even with the falling fundamental frequency pattern the highest number of quantity 3 judgments was 90 out of 104. The peak value for quantity 3 judgments for the whole set of conditions was reached when both conditions were met: the fundamental frequency had a falling pattern and the second syllable was short.

Let me now summarize briefly where we stand with regard to the status of the experiments. I am currently in the process of working out the statistical design for testing the significance of the relationships displayed on this set of tables and figures. I intend to compute correlations between the variables and the judgments and establish the relative contribution of each variable. Until this part of the project is completed, the results are somewhat impressionistic. Nevertheless, it is possible to draw some tentative generalizations.

First of all, I think it is clear that the assignment of a word to a quantity depends not only on the duration of a first syllable vowel or an intervocalic consonant, but also on the duration of the second syllable and on the fundamental frequency pattern applied to the word as a whole. If one defines the point of overlap between two distribution curves as the boundary between two phonemic quantities, one may claim that the placement of these boundaries depends significantly on both second syllable duration and fundamental frequency. I believe that this observation lends support to the

notion that what we are dealing with is a higher-level suprasegmental pattern distributed over the whole disyllabic word, not with independently functioning segmental quantity.

It is interesting, furthermore, that the boundary between quantities 2 and 3 is more strongly affected by the pattern applied to the word as a whole than the boundary between quantities 1 and 2. In a very tentative sense, one might find support here for the idea that the older two-way opposition between short and long is more firmly segmentally anchored than the relatively new three-way opposition between short, long and overlong. The older opposition is mainly segmental; the newer three-way opposition is mainly based on differences between patterns manifested over the whole disyllabic word. The implications of these results will become clearer when the statistical analysis is complete.

## Footnotes

[1]The DDP 24 computer is a machine of medium size (12K) and speed (5 microseconds). The synthesis programs were written by B. E. Caspers (B. E. Caspers, "Software Facilities and Operating System of a DDP- 224 Computer", Bell Telephone Laboratories, Murray Hill, N.J., 1968). I am grateful to Dr. P. B. Denes, Head of the Speech and Communication Research Department, Bell Telephone Laboratories, for his assistance.

[2]I am indebted to Mr. Kullo Vende for his invaluable help in arranging for the listening sessions. I would also like to thank all individuals who participated in the listening tests.

Table 1. Judgments depending on second syllable duration.

| Duration of $V_2$ in msec | taba | tapa | tappa | Total |
|---|---|---|---|---|
| 180 | 784 | 1090 | 310 | 2184 |
| 120 | 686 | 767 | 731 | 2184 |
| 90 | 656 | 731 | 797 | 2184 |
| Total | 2126 | 2588 | 1838 | 6552 |

Table 2. Judgments depending on the duration of /p/

$V_2$ = 180 msec

| Duration of /p/ in msec | taba | tapa | tappa |
|---|---|---|---|
| 40 | 104 | | |
| 50 | 103 | | 1 |
| 60 | 104 | | |
| 70 | 104 | | |
| 80 | 103 | 1 | |
| 90 | 97 | 7 | |
| 100 | 78 | 26 | |
| 110 | 50 | 54 | |
| 120 | 26 | 78 | |
| 130 | 9 | 93 | 2 |
| 140 | 3 | 100 | 1 |
| 150 | 2 | 102 | |
| 160 | | 98 | 6 |
| 170 | | 93 | 11 |
| 180 | | 92 | 12 |
| 190 | 1 | 80 | 23 |
| 200 | | 71 | 33 |
| 210 | | 61 | 43 |
| 220 | | 56 | 48 |
| 230 | | 45 | 59 |
| 240 | | 33 | 71 |
| Total | 784 | 1090 | 310 |

Table 3. Judgments depending on the duration of /p/

$V_2$ = 120 msec

| Duration of /p/ in msec | taba | tapa | tappa |
|---|---|---|---|
| 40 | 104 | | |
| 50 | 103 | 1 | |
| 60 | 102 | 2 | |
| 70 | 99 | 5 | |
| 80 | 96 | 8 | |
| 90 | 83 | 21 | |
| 100 | 58 | 45 | 1 |
| 110 | 33 | 71 | |
| 120 | 4 | 97 | 3 |
| 130 | 1 | 98 | 5 |
| 140 | 3 | 97 | 4 |
| 150 | | 82 | 22 |
| 160 | | 81 | 23 |
| 170 | | 73 | 31 |
| 180 | | 31 | 73 |
| 190 | | 27 | 77 |
| 200 | | 14 | 90 |
| 210 | | 8 | 96 |
| 220 | | 3 | 101 |
| 230 | | 3 | 101 |
| 240 | | | 104 |
| Total | 686 | 767 | 731 |

Table 4. Judgments depending on the duration of /p/

$V_2$ = 90 msec

| Duration of /p/ in msec | taba | tapa | tappa |
|---|---|---|---|
| 40 | 104 | | |
| 50 | 102 | 2 | |
| 60 | 103 | 1 | |
| 70 | 100 | 4 | |
| 80 | 89 | 15 | |
| 90 | 67 | 35 | 2 |
| 100 | 53 | 51 | |
| 110 | 31 | 72 | 1 |
| 120 | 5 | 91 | 8 |
| 130 | | 97 | 7 |
| 140 | | 98 | 6 |
| 150 | | 84 | 20 |
| 160 | 1 | 76 | 27 |
| 170 | | 50 | 54 |
| 180 | | 23 | 81 |
| 190 | 1 | 22 | 81 |
| 200 | | 11 | 93 |
| 210 | | 5 | 99 |
| 220 | | 1 | 103 |
| 230 | | | 104 |
| 240 | | 1 | 103 |
| Total | 656 | 731 | 797 |

Table 5

Judgments depending on second syllable duration (fundamental frequency patterns combined)

| Duration of $V_2$ in msec | sada | saada! | saada | Total |
|---|---|---|---|---|
| 180 | 717 | 1114 | 353 | 2184 |
| 120 | 596 | 1054 | 534 | 2184 |
| 90 | 569 | 942 | 673 | 2184 |
| Total | 1882 | 3110 | 1560 | 6552 |

Judgments depending on fundamental frequency pattern (second syllable durations combined)

| $F_0$ pattern (in Hz) | sada | saada! | saada | Total |
|---|---|---|---|---|
| 120-120/120 | 669 | 1096 | 419 | 2184 |
| 120-120/80 | 605 | 1326 | 253 | 2184 |
| 120-80/80 | 608 | 688 | 888 | 2184 |
| Total | 1882 | 3110 | 1560 | 6552 |

Table 6. Judgments depending on first syllable duration and fundamental frequency pattern (second syllable duration constant at 180 msec)

| $F_0$ pattern (in Hz) | $V_1$ duration (in msec) | sada | saada! | saada | Total |
|---|---|---|---|---|---|
| 120-120/120 | 120 | 101 | 3 | | |
| | 140 | 89 | 15 | | |
| | 160 | 52 | 51 | 1 | |
| | 180 | 17 | 84 | 3 | |
| | 200 | 1 | 93 | 10 | |
| | 220 | | 87 | 17 | |
| | 240 | | 73 | 31 | |
| Total | | 260 | 406 | 62 | 728 |
| 120-120/80 | 120 | 96 | 8 | | |
| | 140 | 85 | 16 | 3 | |
| | 160 | 42 | 57 | 5 | |
| | 180 | 10 | 84 | 10 | |
| | 200 | 3 | 94 | 7 | |
| | 220 | | 89 | 15 | |
| | 240 | 1 | 75 | 28 | |
| Total | | 237 | 423 | 68 | 728 |
| 120-80/80 | 120 | 99 | 5 | | |
| | 140 | 72 | 31 | 1 | |
| | 160 | 41 | 58 | 5 | |
| | 180 | 5 | 78 | 21 | |
| | 200 | 2 | 60 | 42 | |
| | 220 | 1 | 45 | 58 | |
| | 240 | | 8 | 96 | |
| Total | | 220 | 285 | 223 | 728 |
| | | 717 | 1114 | 353 | 2184 |

Table 7. Judgments depending on first syllable duration and fundamental frequency pattern (second syllable duration constant at 120 msec)

| $F_0$ pattern (in Hz) | $V_1$ duration (in msec) | sada | sada! | sada | Total |
|---|---|---|---|---|---|
| 120-120/120 | 120 | 95 | 8 | 1 | |
| | 140 | 77 | 27 | | |
| | 160 | 23 | 72 | 9 | |
| | 180 | 10 | 82 | 12 | |
| | 200 | 2 | 77 | 25 | |
| | 220 | 1 | 61 | 42 | |
| | 240 | 1 | 34 | 69 | |
| Total | | 209 | 361 | 158 | 728 |
| 120-120/80 | 120 | 96 | 8 | | |
| | 140 | 78 | 25 | 1 | |
| | 160 | 17 | 83 | 4 | |
| | 180 | 7 | 90 | 7 | |
| | 200 | | 92 | 12 | |
| | 220 | | 92 | 12 | |
| | 240 | 1 | 70 | 33 | |
| Total | | 199 | 460 | 69 | 728 |
| 120-80/80 | 120 | 87 | 15 | 2 | |
| | 140 | 69 | 33 | 2 | |
| | 160 | 17 | 73 | 12 | |
| | 180 | 10 | 58 | 36 | |
| | 200 | 1 | 27 | 76 | |
| | 220 | 3 | 12 | 89 | |
| | 240 | 1 | 13 | 90 | |
| Total | | 188 | 233 | 307 | 728 |
| | | 596 | 1054 | 534 | 2184 |

Table 8. Judgments depending on first syllable duration and fundamental frequency pattern (second syllable duration constant at 90 msec)

| $F_0$ pattern (in Hz) | $V_1$ duration (in msec) | sada | saada! | saada | Total |
|---|---|---|---|---|---|
| 120-120/120 | 120 | 74 | 26 | 4 | |
| | 140 | 76 | 27 | 1 | |
| | 160 | 32 | 63 | 9 | |
| | 180 | 14 | 77 | 13 | |
| | 200 | 3 | 68 | 33 | |
| | 220 | 1 | 40 | 63 | |
| | 240 | | 28 | 76 | |
| Total | | 200 | 329 | 199 | 728 |
| 120-120/80 | 120 | 78 | 25 | 1 | |
| | 140 | 58 | 44 | 2 | |
| | 160 | 22 | 78 | 4 | |
| | 180 | 9 | 86 | 9 | |
| | 200 | | 87 | 17 | |
| | 220 | 1 | 69 | 34 | |
| | 240 | 1 | 54 | 49 | |
| Total | | 169 | 443 | 116 | 728 |
| 120-80/80 | 120 | 87 | 17 | | |
| | 140 | 79 | 19 | 6 | |
| | 160 | 15 | 64 | 25 | |
| | 180 | 14 | 37 | 53 | |
| | 200 | 1 | 20 | 83 | |
| | 220 | 2 | 8 | 94 | |
| | 240 | 2 | 5 | 97 | |
| Total | | 200 | 170 | 358 | 728 |
| | | 569 | 942 | 673 | 2184 |

Figure 1. Number of judgments as taba, tapa or tappa, expressed as a function of the duration of the second syllable.

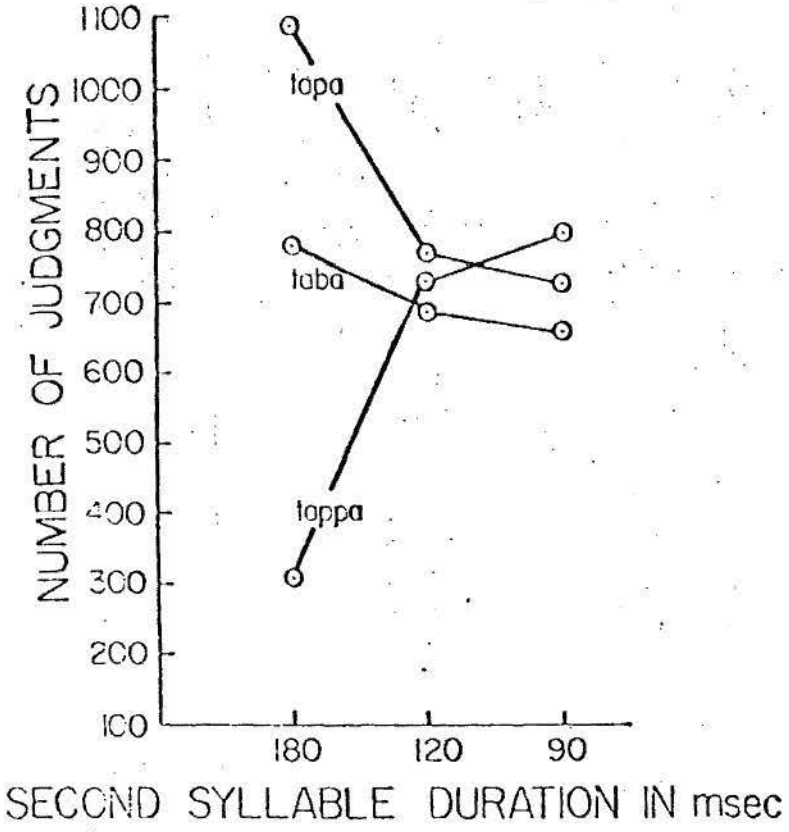


Figure 2. Number of judgments as taba, tapa or tappa, expressed as a function of the duration of intervocalic /p/. Duration of the second syllable was constant at 180 msec.

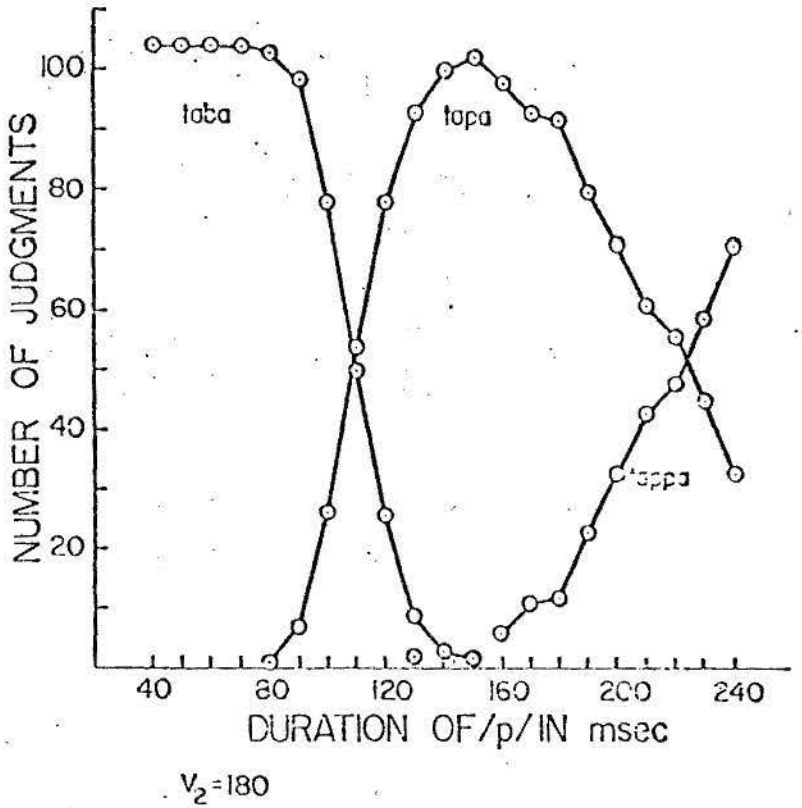

Figure 3. Number of judgments as taba, tapa or tappa, expressed as a function of the duration of intervocalic /p/. Duration of the second syllable was constant at 120 msec.

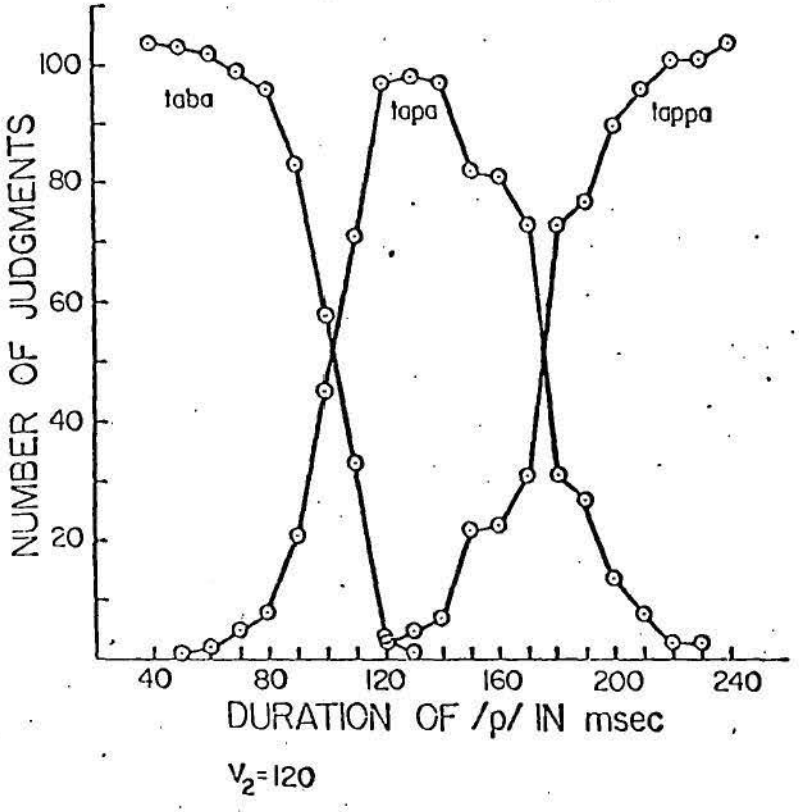


Figure 4. Number of judgments as taba, tapa or tappa, expressed as a function of the duration of intervocalic /p/. Duration of the second syllable was constant at 90 msec.

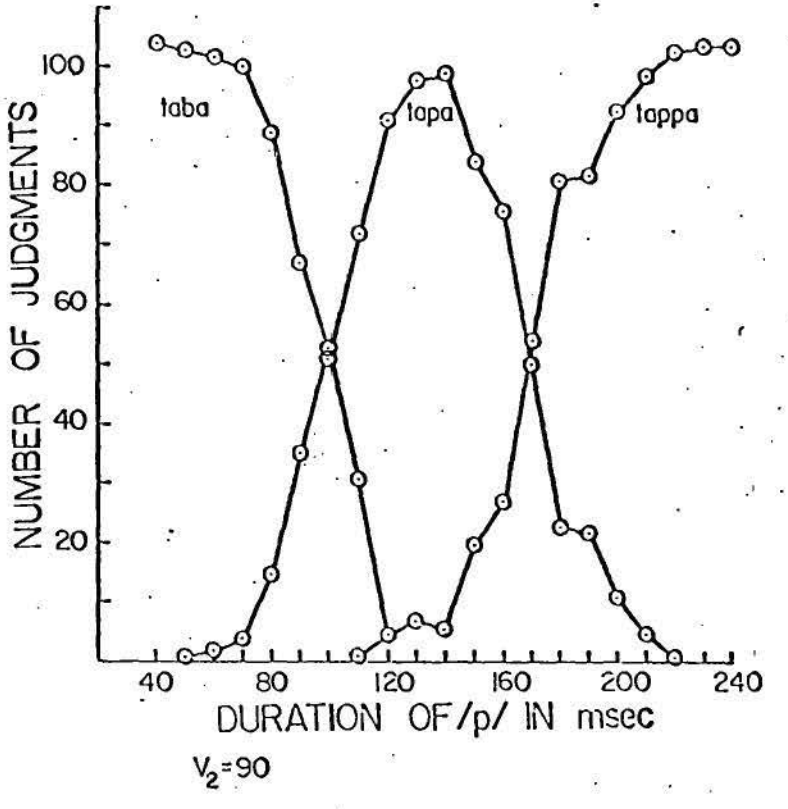

Figure 5. Number of judgments as sada, saada! or saada, expressed as a function of the duration of the second syllable (with first syllable duration and fundamental frequency patterns combined) and as a function of fundamental frequency pattern (with first and second syllable durations combined). Fundamental frequencies are given in Hz.

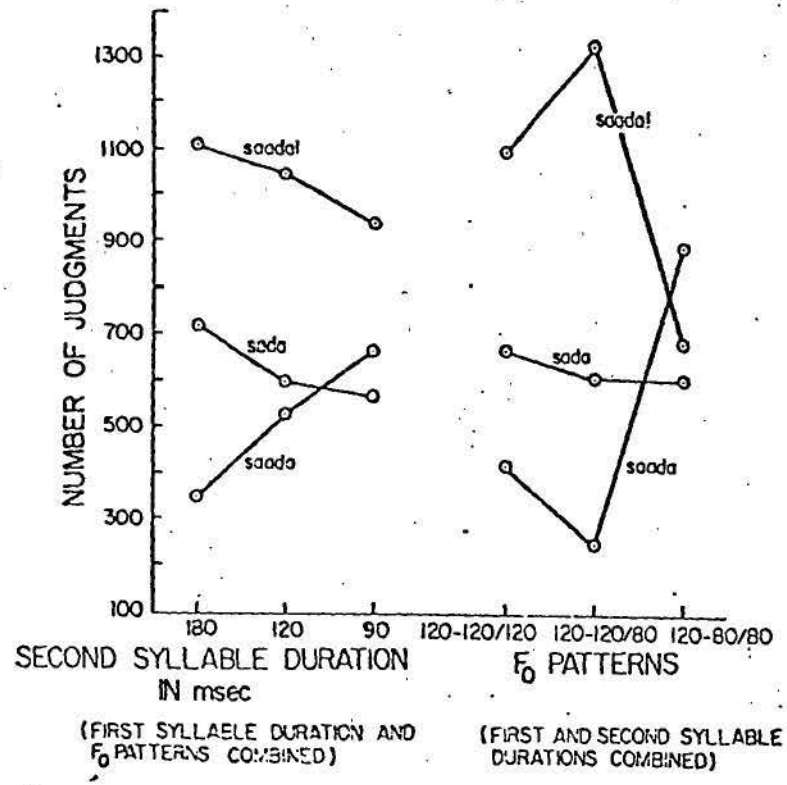


Figure 6. Number of judgments as sada, saada! or saada, expressed as a function of the duration of the first syllable. The duration of the second syllable was 180 msec, the fundamental frequency pattern was level at 120 Hz.

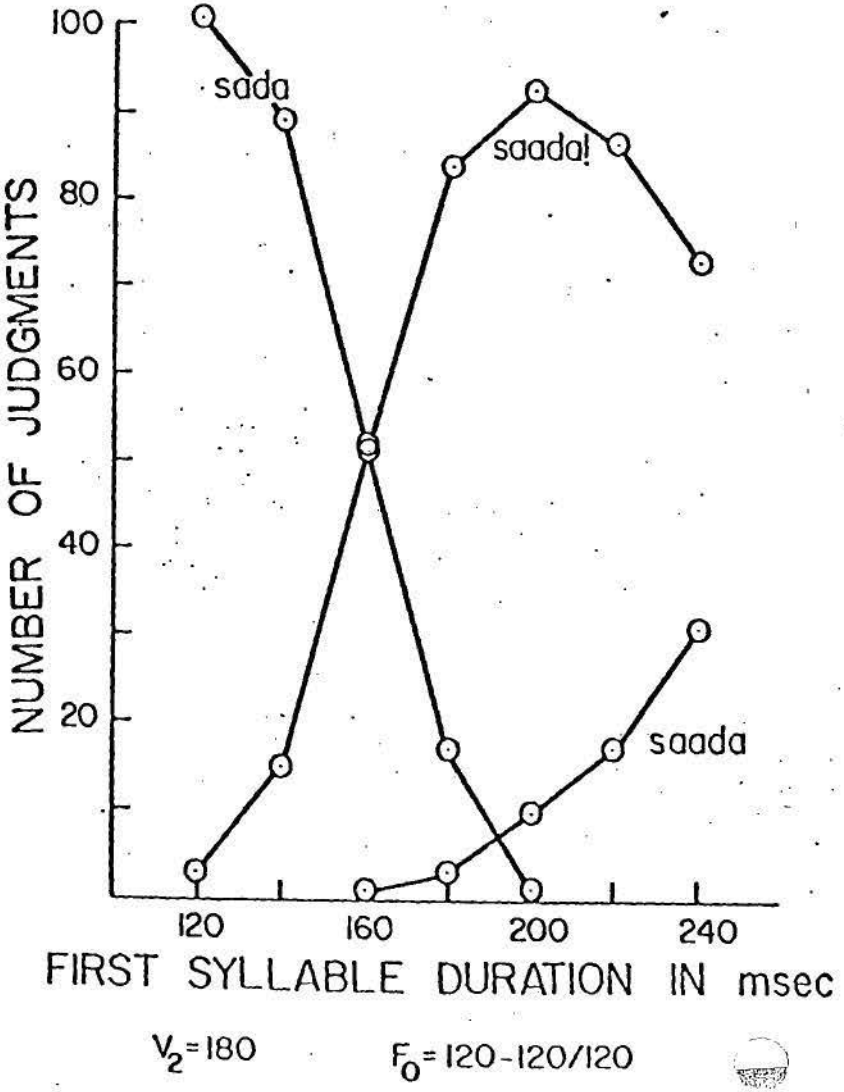

Figure 7. Number of judgments as sada, saada! saada, expressed as a function of the duration of the first syllable. The duration of the second syllable was 180 msec, the fundamental frequency pattern was falling during the first syllable.

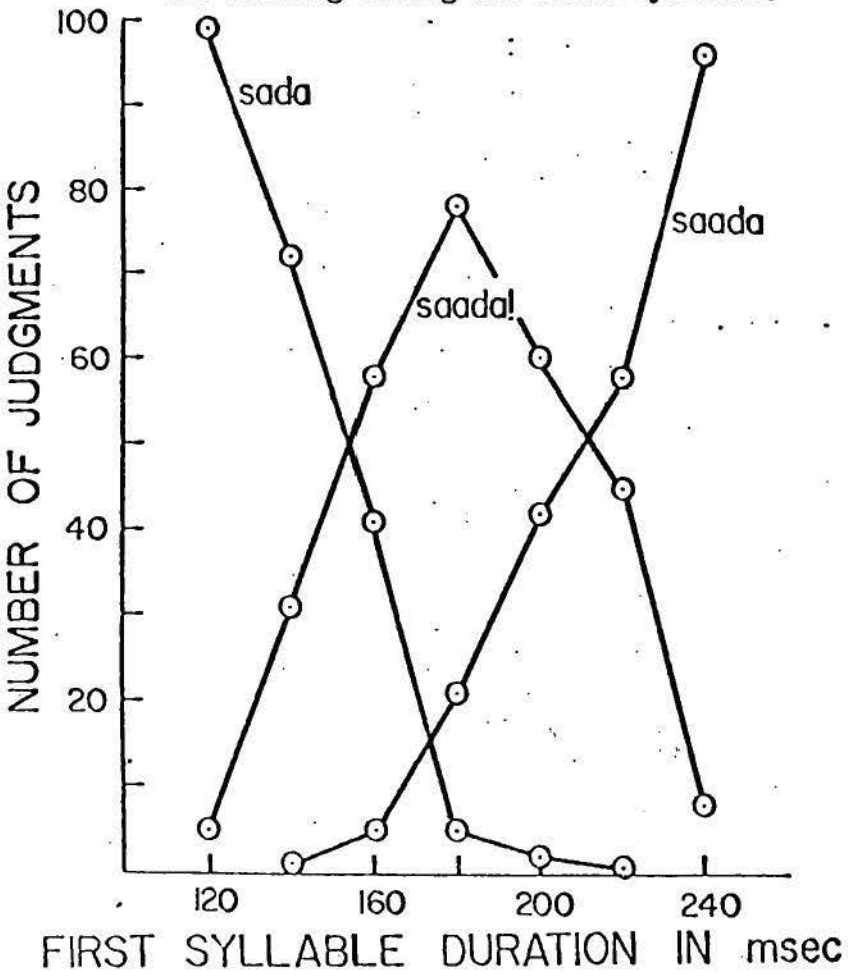


Figure 8. Number of judgments as sada, saada! or saada, expressed as a function of the duration of the first syllable. The duration of the second syllable was 90 msec, the fundamental frequency pattern was level at 120 Hz.

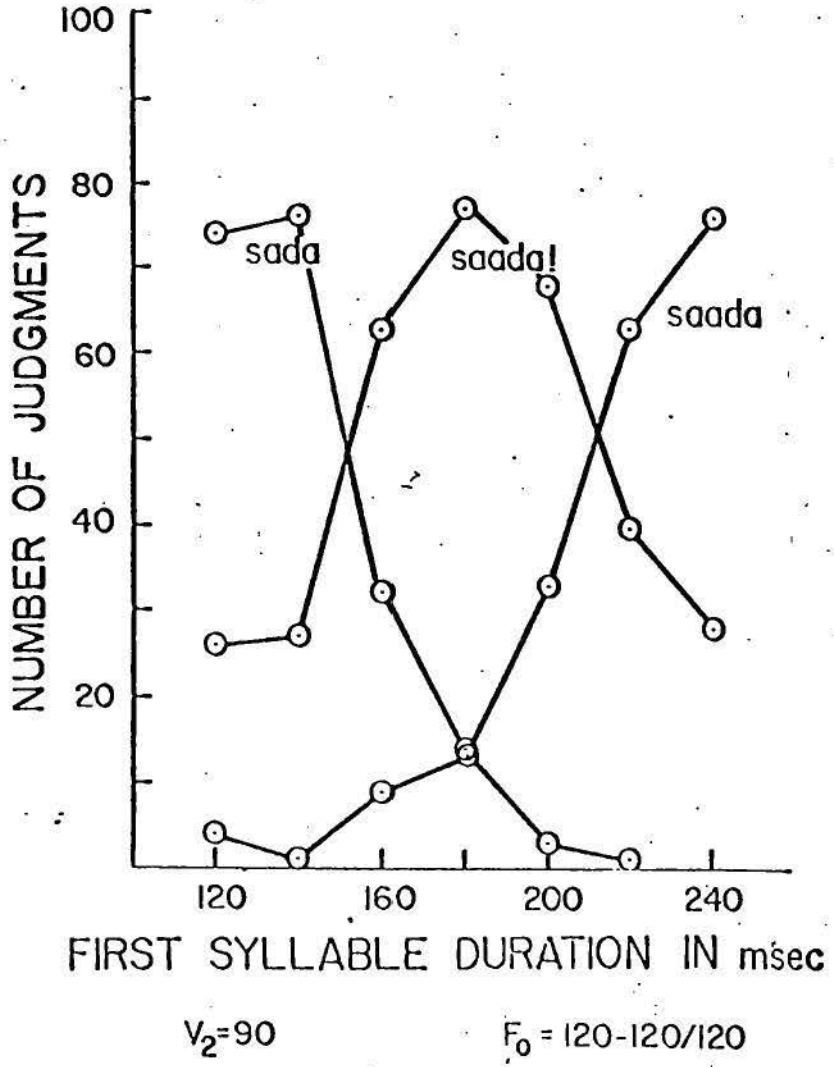

Figure 9. Number of judgments as sada, saada! or saada, expressed as a function of the duration of the first syllable. The duration of the second syllable was 90 msec, the fundamental frequency pattern was falling during the first syllable.

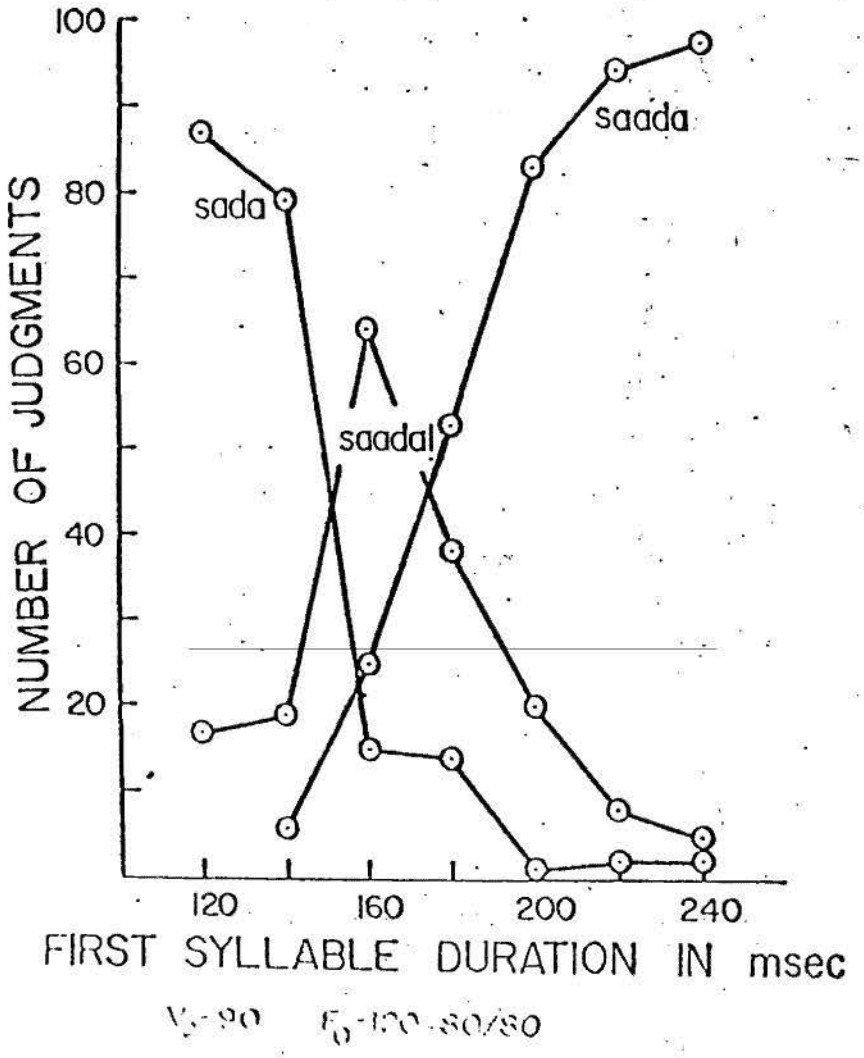


Figure 10. Number of judgments as saada! (quantity 2), expressed as a function of the duration of the first syllable and the fundamental frequency pattern.

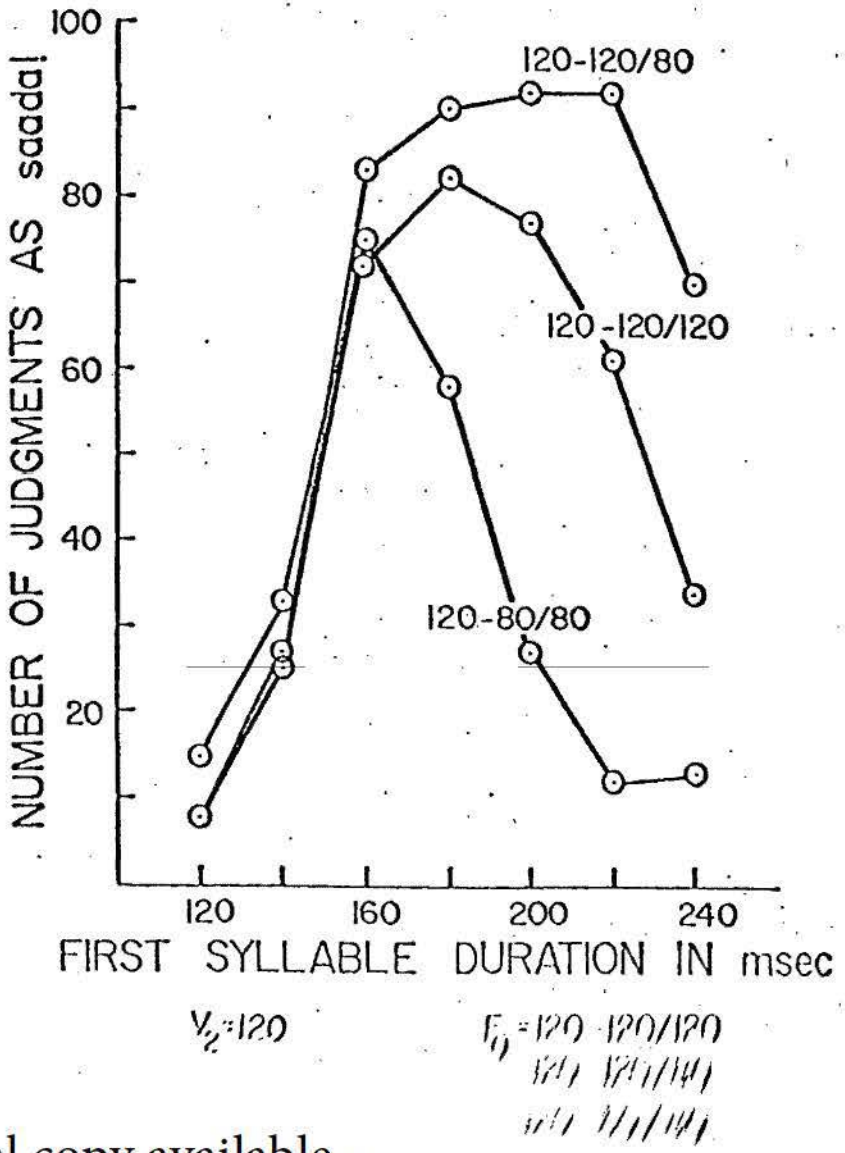

Figure 11. Number of judgments as saada (quantity 3), expressed as a function of the duration of the first syllable and the fundamental frequency pattern.